Downloaded from orbit.dtu.dk on: May 06, 2023

DTU Library

Forskningsdata og Open Access

et Deff-Projekt

Heller, Alfred; Blaabjerg, Niels Jørgen; Clausen, Nanna Floor; Christensen-Dalsgaard, Birte; Dorch, Bertil

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Heller, A., Blaabjerg, N. J., Clausen, N. F., Christensen-Dalsgaard, B., & Dorch, B. (2011). *Forskningsdata og Open Access: et Deff-Projekt*. DEFF.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Forskningsdata og Open Access

et Deff-projekt

Forfattere:

Alfred Heller, Danmarks Teknisk Informationscenter, DTU (Projektleder)

Niels Jørgen Blaabjerg, Ålborg Universitetsbibliotek

Nanna Floor Clausen, Dansk Data Arkiv

Birte Christensen-Dalsgaard, Kongelig Bibliotek

Bertil Dorch, KUBIS, Københavns Universitet

1.1 Indholdsfortegnelse

Forskningsdata og Open Access

Hvad er relevant for danske forskningsbiblioteker?

2 Forord

Titlen i projektet knytter tydelig en sammenhæng mellem forskningsdata og Open Access (OA) for publikationsområdet. Det skyldes nok mest de erfaringer der stammer fra publiceringsområdet, hvor omkostningsstrukturen bag licensbelagte forskningsresultater har medført problemstillinger. Samme problemstillinger vil man helst undgå for forskningsdata.

Forudsætningerne for forskningsdata er helt forskellige fra publikationer, hvilket en Knowledge Exchange rapport vil kunne dokumentere i løbet af sommeren 2011. Rapporten vil kunne downloades fra http://www.knowledge-exchange.info. Rapporten om de juridiske vilkår for forskningsdata vil dokumentere, at man som udgangspunkt ikke kan ophavsretbeskytte (ikke-personlige) rådata, samt andre data, hvor der ikke indgår betragtelige investeringer i selve databasearkitekturen (se nedenfor), idet data anses som repræsentationer af fakta. Dette gælder for rigtig mange forskningsdata.

Personlige data er beskyttede data og skal "anonymiseres" før publiceringer og deling. Også dette virker ikke som hindring for åben adgang til data.

En anden forskel for forskningsdata i forhold til publiceringer er, at forlagene ikke ønsker at drive store datasamlinger, da dette er meget krævende.

Hvis nærværende rapport underspiller OA-aspektet, er det ud fra den betragtning, at man i de fleste tilfælde kan antage, at forskningsdata i modsætning til skriftlige arbejder ikke er beskyttet af ophavsret. Det er først ved datasamlinger oparbejdet via "essentielt store investeringer"[1], der giver forlagene mulighed for at licensbelægge forskningsdata. Dette er konsekvensen af det europæiske databasedirektiv, og dertil skal lægges, at det kun er investeringer i databasearkitekturen, der skal medtages, når investeringen skal vurderes - ikke investeringer i dataindsamlingen.

Da nærværende projekt blev initieret, var bevidsthedsniveauet og viden omkring forskningsdata begrænset til enkelte personer og få institutioner. I den forløbende periode blev bevidsthedsgraden og interessen afgørende øget, hvilket medfører at viden i nærværende rapport og tilhørende wiki vil forventes at have stort interesse i forskningsorganisationer. Projektgruppen ser denne udvikling som yderst positiv og ser, at de vanskeligheder vi har oplevet undervejs vil komme til at løses op af den øgede bevidsthed og viljen til at sikre og dele forskningsdata.

[1] Kravet om"essentielt store investeringer" fremføres af den kommende KE-rapport "The legal status of research data in Knowledge Exchange partner countries", Madeleine de Cock Buning et.al., Centre for Intellectual Property Law (CIER), Molengraaff Institute for Private Law, Utrecht University, The Netherlands,(2011).

3 Indledning

Projektet "Forskningsdata og Open Access", DEFF:S.Nr. 2010-004751, har kørt i 2010 med forlængelse indtil april 2011. Nærværende rapport har til formål at videndele den indsigt der er skabt i projektet.

Projektet har udover denne rapport genereret en wiki på http://forskningsdata.deff.wikispaces.net – som er mest omfattende afslutningsrapport og dokumentation for projektet.

Deltagere i projektet var DTIC, AUB, KB, KUBIS i samarbejde med Dansk Data Arkiv (DDA). Projektet blev fulgt af en referencegruppe med deltager som i høj grad er sammenfaldende med DEFF-programgruppe for Informationsforsyning.

Projektets baggrund, formål og motivation er hentet fra ansøgningen:

3.1 Baggrund og motivation

Forskningsprojekter producerer og indsamler i stærkt stigende omfang digitale data, der udgør grundlaget for forskernes egne konklusioner og publikationer såvel som grundlaget for de efterfølgende forskningsprojekter – i det omfang, de kan få adgang til data.

Problemerne omkring adgang til og organisering af forskningsdata indgår ofte som del af den såkaldte eScience-diskussion, der bredt ser på udfordringer og muligheder i forbindelse med digital og netværksbaseret forskning. I sit høringssvar[i] til Nordforsks "Nordic eScience Action Plan" skriver DEFF:

> *DEFF finder det essentielt, at forskningsbibliotekernes vitale rolle i forskningens og de højere uddannelsers infrastruktur udnyttes optimalt i forbindelse med planlægning af eScience. Hvordan begrebet eScience end anskues, ender vi i sidste instans med store sæt af primære forskningsdata som skal behandles som bibliotekariske objekter; dvs. objekter med behov for beskrivelse til brug for organisation og genfinding, om det så skal ske automatisk eller intellektuelt. Integration og tilgængeliggørelse af store datasæt med henblik på identifikation, genfinding, udveksling og genbrug fordrer beskrivelse, organisation og ensartet kontrol og behandling ud fra harmoniserede retningslinjer, formater og protokoller. Det er præcis dét, forskningsbibliotekerne gør i dag med dokumenter.*
>
> *Der er den store forskel på data og publikationer, at sidstnævnte gennemgår en overskuelig og forudsigelig proces i forbindelse med publicering, som harmoniserer outputtet og kaster en række sammenlignelige og velkendte datakategorier af sig (fx titel og udgiver). Data, derimod, er en ganske anderledes størrelse. For data er der intet organisationelt system på plads, der kan sammenlignes med publikationsprocessen og den efterfølgende bibliografiske kontrol. Det gør data meget domænespecifikke, kontekstuelt forankrede og ganske enkelt dårligt sammenlignelige fra datasæt til datasæt. Datasæt tabes hver dag for stedse af denne årsag alene, og afskrivningen af anvendte ressourcer i forbindelse med tabte eller ikke-genfindelige data er betragtelige. Lykkeligvis kan meget af det intellektuelle, bibliotekariske arbejde med dokumenter ekstrapoleres til arbejdet med data.*

Siden disse linjer er skrevet, blev behovet for arbejdet omkring forskningsdata yderligere aktualiseret. Det er Deff's opgave at placere forskningsdata i den kommende e-Science Center som etableres gennem en infrastrukturpulje i sammenarbejde med Danish Scientific Computation Center og Forskningsnettet. Modsætning må være at etablere e-Science inkluderende forskningsdata, lignende den organisation man har valgt i de australske, hollandske og svenske organisationer af samme karakter.

Fra DEFF Programgruppen for Informationsforsynings handlingsplan 2009-10:

> ***C: Forskningsdata og Open Access***
> *– med følgende målsætninger for perioden:*

1. *At gennemføre pilotprojekt(er), der i praksis skal bidrage til en afklaring af bibliotekernes rolle ifm.* ***repositories for forskningsdata****. Der lægges vægt på Open Access, på samarbejde med eksisterende datacentre og med aktiviteterne i Knowledge Exchange "Primary Research Data" samt evt. kommende NordForsk eScience initiativer.*
2. *At gennemføre projekt omkring persistent* ***linkning fra publikationer til forskningsdata.*** *Der lægges vægt på en afklaring af bibliotekernes bidrag samt på samarbejde med datacentre og med aktiviteterne i Knowledge Exchange "Primary Research Data" samt evt. kommende NordForsk eScience initiativer.*

3.2 Projektresumé og formål

Pilotprojektet skal bidrage til en afklaring af bibliotekernes rolle i forhold til forskningsdata.

Pilotprojektet skal støtte DEFF's fremtidige planlægning af open-access til forskningsdata …

Projektet vil levere input til de strategiske overvejelser omkring DEFF og bibliotekernes position og involvering i forskningsdataproblematikken. Nogle overordnede spørgsmål der bør afklares er:

- Hvilke muligheder og udfordringer findes på dette område?
- Hvilke indsatsområder er relevante for bibliotekerne?

Projektet leverer input til den strategisk proces ved at undersøge følgende spørgsmål:

- Hvilke interessenter/aktører er involveret i forskningsdata? Hvad er deres rolle?
- Hvordan kan bibliotekerne fremstå som relevante og interessante partnerne på dataområdet?
- Hvorledes sikres sammenhæng mellem publikationer og forskningsdata?
- Hvilke erfaringer findes internationalt på emnet?

Herudover er der formuleret en masse spørgsmål, som projektet havde til formål, at undersøge og give svar på. Svarene gives i det følgende.

4 Metode og leverancer

I ansøgningen var følgende fremgangsmåde planlagt:

4.1 Projektets indsatser og mål

Pilotprojekt adresserer de nævnte målsætninger gennem studier af den internationale "best practice" og ved gennemførelse af overkommelige lokale eksperimenter med det formål, at projektet leverer de første bidrag til en afklaring af DEFF's og bibliotekernes rolle på de to nævnte områder:

1. Modtagelse, organisering og opbevaring af digitale forskningsdata med henblik på adgang via nettet i et langsigtet Open Access-perspektiv.
2. Sammenkobling af forskningspublikationer og de datasæt, de bygger på, således at der kan linkes konsekvent og persistent fra digitale forskningspublikationer til digitale forskningsdata.

Pilotprojektet vil mere specifikt se på om institutionernes forskningsdatabaser og Institutional Repositories, der i dag er gode bud på håndtering af de digitale publikationer, men (endnu) ikke spiller en rolle i forbindelse med forskningsdata, vil kunne anvendes til formålet.

Det nødvendige samarbejde mellem forskere og biblioteker på området findes heller ikke i dag, hvorfor det vil være en udfordring for begge parter at få dette etableret. Mens der er tale om en helt ny udfordring for bibliotekarerne, skal forskerne overbevises om, at bibliotekerne er i stand til at håndtere deres data i henhold til deres interesser, f.eks. embargoperioder og lignende. Derfor vil nærværende projekt i praksis også fokusere på de organisatoriske udfordringer.

Konkret vil projektet levere en eller flere demonstratorer, hvor der kan linkes fra en publikationsbeskrivelse i en forskningsdatabase/Institutional Repository til det/de relevante datasæt – f.eks. som regneark, hvis datamængde og -struktur tillader dette.

Herudover vil projektet benytte, teste og evaluere den nye DataCite infrastruktur, der netop er etableret i et bredt internationalt samarbejde mellem videnskabelige informationscentre og biblioteker[ii]. Med DataCite (International Initiative to Facilitate Access to Research Data) tildeles videnskabelige datasæt den samme slags identifikatorer (DOI'er), som forlagene tildeler publikationer gennem Crossref. Herved vil man kunne linke lige så let og sikkert fra en publikation til dens datasæt som fra en publikation til dens citerede publikationer. Disse kvalitetslinks vil kunne optræde såvel i selve publikationerne som i metadata-beskrivelser af publikationerne, således som det vil være tilfældet i pilotprojektets danske forskningsdatabaser og/eller Institutional Repositories.

Der arbejdes internationalt med publicering af forskningsdata, da området stadig har udfordringer at løse. Projektet vil kunne hente (og videreformidle) væsentlige erfaringer fra Knowledge Exchange's "Primary Research Data"-aktiviteter[iii], der bl.a. omfatter en Working Group. Ligeledes har KE-partneren JISC i de senere år gennemført en række projekter på området, der bør kunne levere input til "best practice overview". Det kan fremhæves, at deltagerne i projektgruppen allerede er involveret i disse internationale aktiviteter, hvilket projektet vil drage nytte af.

Det internationale perspektiv blev inddraget i arbejdet, gennem deltagelse i Knowledge Exchange, DataCite, IDF. Også blev det til en studietur til Australien, som har givet projektdeltagere og referencegruppen en afgørende indsigt i forskningsdataområde. Mere om det i afsnittet om "Best Practice" i denne rapport.

5 Resultater, diskussioner og konklusioner

Projektet har medført mange resultater som er forsøgt at gengive i Bilag 1 i kort form. I nærværende afsnit forsøges de vigtigste gengivet i detaljer.

5.1 Strategiarbejdet som del af projektet

Projektet var aktiv omkring strategiarbejdet i de første 3-4 måneder af projektet, da Deff-strategien blev lagt i denne periode. Dette skete ved at opsamle de erfaringer er lå til rådighed af foregående projekter og internationale erfaringer. Disse input til strategien blev skrevet som forslag til Deff-styregruppen og Claus Vestager Pedersen og Alfred Heller deltog ved et strategiworkshop med styregruppen i forår 2010 for at svare på spørgsmål fra styregruppen. Resultatet heraf kan aflæses af Deff-strategien som forventes udgivet i efterår 2011.

5.2 Aktører og roller

Forskningsdata er et arbejdsområde som bibliotekerne ikke anser selv eller af deres omgivelser for at være helt oplagt som indsatsfelt. Motiveret af udenlandske succeshistorier, forsøgte projektets deltager at finde samarbejder mellem bibliotekerne og forskerne.

I projektet blev der gennemført flere forsøg på at etablere samarbejde omkring opsamling og bevaring af data – her nogle konkrete erfaringer:

Vejrdata: En forsker havde opsamlet vejrdata gennem sin lange karriere. Han var utrolig påpasselig med at sikre data for hele perioden og han overført dem igennem tiden til de medier der var tilgængelige, fra hulkort til disks og tapes til CD-rommer. Dataeksempler blev lægt ind i Fedora-Commons prototypen, hvor de blev beskrevet i DataCite og Dublin Core formaterne, samt etableret forklarende relationer mellem de forskellige datafiler. Formålet var at skabe en sammenhængende struktur af relationer for at kunne etablere en service, der

kunne levere ønskede datasekvenser og –perioder. Det endte alligevel med, at forskeren gik på pension, før hans data blev sikret. Instituttet der overtog hans arbejde og skal videreføre målingerne fokuserer kun på de nye data og det nye udstyr. Historien kan findes andre steder. Der er en vilje på institutterne til at samarbejde, men den enkelte forsker har ikke kapacitet til at udføre det meget omfattende arbejde med at finde data, beskrive data tilstrækkelig osv.

I en anden case kom data fra interviews i kombination med nogle målinger af indeklima i bygninger. Data blev lagt ind i samme system som forrige case og der blev etableret de nødvendige metadata til formålet manuel. Tilgangen til dataene blev etableret for forskerne selv, så de kunne arbejde videre på denne platform. Tilgangen til udefrakommende blev stoppet, da en anden forsker gjorde krav på at dataene skulle holdes lukket af forskellige grunde. Derfor kunne disse data ikke skabe grundlag for en demonstrator.

På det organisatoriske niveau blev der ikke stillet spørgsmålstegn ved, at bibliotekerne involverer sig i dette arbejde. Forskerne er klar over at bibliotekerne har tradition for at opbevare forskningspubliceringer over lang tid og uden tab. Skepsis består dog i forhold til den IT-tekniske kompetence, hvor man umiddelbar ville lade IT-afdelinger tage sig af dette arbejde. Forskningsbiblioteker der har etablerede IT-udviklingsafdelinger, er dette dog meget hurtig forklaret og giver ikke anledning til modstand. I andre tilfælde er et samarbejde mellem biblioteker og IT-afdelinger en nødvendighed.

Der blev taget kontakt til et stort tal forskere og forskergrupper som alle viser behov for hjælp. Store projekter har data management planer og etablerede IT-infrastukturer – ofte dog lukkede bag ved projekternes eller organisationernes interne systemer.

Det var mange grunde til at forskerne ikke ville indgå i projekter med forskningsbibliotekerne, hvor manglende ressourcer er det største grund til ikke at gøre det. Angst om rettigheder eller fejlfortolkning af tilgængelige data var/er en anden nævnt grund.

Bibliotekarerne er meget interesserede i forskningsdata som sådan og kan se det som en udvidelse af forskerservices, som er ved at være et etableret emne i faggruppen. Når det så kommer til stykket ser man følgende mønstre:

Selv en meget erfaren bibliotekar der kender til samarbejde med forskerne, systemer, forskningsregistrering, metadata, IT osv. har vist yderst stor besvær med at sætte sig ind i forskningsdataområde. Det er simpelthen et fremmede stof og hver case har sit helt eget indhold, baggrund, formål og dermed håndtering. (Anvendelse af en software til håndtering, som DataVerse og Fedore-Commons i en afrundet version til formålet, ville skabe ro og tryghed.)

Unge studentermedhjælpere på biblioteker, som ellers arbejder med bl.a. forskningsregistrering og derigennem kender til metadata, viste ikke den ovenfor nævnte "besvær" med at tilgå forskningsdata. Det var lige til at lægge disse data ind i systemet, beskrive dem med metadata, mens de hurtig kunne oplære konceptet af relationer mellem dataklumperne. Dette viser ret tydelig at

forsknings- og systembibliotekarer har forudsætningen til at udføre arbejdet, og at en vis tilpasning er nødvendigt for at få dem til det.

5.3 Aktører og roller – Organisationer

Hvordan står forskningsbibliotekerne i forhold til andre aktører til håndtering af forskningsdata? Forskerne kender biblioteker som troværdige organisationer som man arbejder med dagligt. Forskere er vant til at arbejde med biblioteker, omvendt kender de ofte ikke til de brede tilbud der i virkeligheden eksisterer på forskningsbibliotekerne. Alt i alt står forskningsbiblioteker godt til at løse denne opgaver. Ser man forholdet til andre organisationer om IT, så skal der oftest findes et samarbejde mellem lokale aktører og også med nationale aktører.

Forskningsbibliotekerne som organisation og ledelserne på disse har naturligvis en afgørende rolle for en accept af forskningsdataområde. Det opleves, at erfarne bibliotekarer har større besvær med at sætte sig ind i emnet i forhold til studentermedhjælpere, hvilket ikke kun skyldes personlige forhold. Dette skyldes måske også den omstændighed, at "ledelsen" på forskningsbibliotekerne ikke har meldt ud, at arbejdsfeltet anses som en mulig karrierevej. Australske erfaringer peger på lige præcis dette: På Monash Universitey i Melbourne, var det selveste biblioteksdirektøren der satte sig først ind i stoffet, hvorefter hun oplærte de første bibliotekarer i det. Dermed blev der sat et fokus på området, som medførte at emnet blev modtaget på en positiv måde.

Andre organisationer så som IT-organisationer og arkiver ville være naturlige aktører på området. Det viser sig dog, at disse ikke er interesseret i samme aspekter som bibliotekerne: IT-organisationer vil typisk stille hardware og forbindelser til rådighed, mens arkiverne vil "arkivere" nogle ikke offentlige versioner til side. Ingen af disse organisationer synes interesserede med at gøre informationer tilgængelige og genbrugelige eller langtidsholdbare i en aktuel og brugbar version. En undtagelse er Dansk Data Arkiv, der har erfaring med både at langtidsbevare og dokumentere forskningsdata, men, som ligesom bibliotekerne, har den samme interesse i udlån af data. Dansk Data Arkiv er pt. ved at udvide sit område indenfor forskningsdata og et tættere samarbejde mellem forskningsbibliotekerne og Dansk Data Arkiv vil være et godt udbytte af dette projekt.

De nationale strukturer og forskningsfinansierende organisationer spiller en afgørende rolle for fremtiden af forskningsdata. I Denmark synes de involverede myndigheder ikke at være klar over behovet for håndtering og bevaring af forskningsdata. Senest blev dette underbygget ved, at en national e-Science-infrastruktur under infratrukturpuljen, ikke har medtaget en konkret "pind" omkring forskningsdata. Som med open acces, er det bibliotekerne der sætter emnet på dagsordenen. Da emnet er så enormt økonomikrævende, er det ikke muligt for forskningsbibliotekerne at gennemføre projektet. Derfor skal de nationale finansieringskilder og fonder til at løfte området.
Forskningsbibliotekerne vil på dette tidspunkt forhåbentlig være på pletten, når man skal etablere disse infrastrukturer og ses som relevant organisation hertil, mens andre har sovet i timerne. DEFF er en naturlig bannerfører for området.

5.4 Autogenerering af metadata for forskningsdata

KB har behandlet to cases, en omkring arkæologiske data i samarbejde med Institut for tværkulturelle og regionale studier og et i samarbejde med Astronomisk Institut i Århus.

5.4.1 Case: Arkæologiske data

KB's case tager fat i arkæologerne på ToRS's konkrete udgravningsprojekt i Jordan. Institut for Tværkulturelle og Regionale Studier (ToRS) ved Københavns Universitet og KB samarbejder om projektet 'Arkæologisk VRE'. Udgangspunktet for projektet er et ønske fra ToRS om, at der stilles et system til rådighed, der hoster, videreformidler og opbevarer videnskabelige digitale primærdata (rå data) til brug for forskere ved ToRS, det øvrige KU, andre forskningsinstitutioner og arkæologi-interesserede medlemmer af den brede offentlighed.

Formål:

Formålet med projektet er, at beskrive og opbygge et sådant modulært system i samarbejde med forskerne fra ToRs, således at de får den service, der opfylder deres ønske: At bevare og sikre konkrete arkæologiske primære data, at fascilitere samarbejde på tværs af institutioner og lande om disse data og at give forskningen et ansigt udadtil mod offentligheden ved at formidle udvalgte data og heraf afledte publikationer.

Systemet foreslås baseret på moduler, fx Harvard Dataverse (http://dvn.iq.harvard.edu/dvn/), Hprints (http://www.hprints.org/) og Sharepoint 2007 hvor 1) Dataverse indgår som repository for digitale primærdata, 2) Hprints som (Open Access) repository for publikationer, artikler mv. og 3) Sharepoint som præsentationslag (CMS).

Primærdata:

Arkæologernes primær data består af:

· Locusark (udgravningsregistrering, hidtil på papir)

· Plantegninger (håndtegnede, hidtil på papir)

· Profiltegninger (håndtegnede, hidtil på papir)

· Objekt- og flinttegn (håndtegnede, hidtil på papir)

· Rekonstruktioner (håndtegnede, hidtil på papir)

· Analyseresultater, C-14 årstalsdateringer

· Fotomateriale (født digitale fotos) af fundsteder, objekter, landskaber

Automatisk generering af metadata:

Omkring digitale fotos og C-14 er der potentiale i at indsamle og generere metadata on-the-fly. For C-14 dateringer vil metadata centrere sig om det pågældende laboratoriums test-metode og test-udstyr, hvilket kan variere idet der anvendes laboratorier i både Europa og New Zealand (man køber sig til

ydelser fra laboratorier der byder ind med den billigste pris, hvorfor man løbende skifter laboratorium). For digitale fotos er det allerede praksis at dagens digitale kameraers EXIF-data (Exchangeable Image File Format) har metadata indlejret. EXIF er en branche de-facto standard, men er ikke vedligeholdt af en standardiserings-organisation. For digitale fotos vil det af langtidsbevarings-hensyn være relevant at metadata genereres for objekterne, for at sikre at bevarede fotos kan porteres til nye fremtidige standarder/teknologier.

Kravspecifikation:

Kravspecifikation bliver udarbejdet på baggrund af interview med ToRS repræsentanter. Der pågår p.t. et arbejde ifht.. at indsamle relevante data fra ToRS' hidtidige repositorier (filemaker-database), fotos, tegninger, Locus-ark mv. Kravspecifikationen vil bl.a. indeholde en gennemgang af automatisk metadata-generering ifht. fotos og C-14 prøver.

5.4.2 Case: Astroseismiske data

Indledning: Forskningsdata - i sammenhæng

Sammenkædning af data og artikler afspejler en meget udbredt, men ikke desto mindre traditionel tankegang i forhold til kommunikation om erkendelse og forståelse af ny viden. Data er basis for en analyse, som efterfølgende publiceres, meget gerne med direkte reference til disse bagvedliggende datasæt og med reference til analysemetoden. Artiklen er i sig selv data. Som kan bruges i videre forskning – Nature "Speed Reading" – hvilket ofte kræver at man finder, læser og evt. får adgang til data.

Interessant nok er der flere, som er begyndt for alvor at stille spørgsmålstegn ved den nuværende strategi og peger på at nye metoder, som understøtter et anderledes overblik og analyse, er nødvendigt. Der er også flere, som er begyndt at pege på nødvendigheden af de såkaldte implicitte informationer (tacid information), som er nødvendige for at forskere udenfor ens eget snævre felt, kan bruge data – dvs. forstå, hvordan de er skabt og hvilke forhold potentielt kan påvirke resultaterne. Man er også nødt til at forklare, hvordan en evt. dataoprensning og efterbehandling er foregået. Et eksempel er beskrevet i en review artikel i Biochemical Journal, "Calling International Rescue: knowledge lost in literature and data landslide!", som kan ses på http://www.biochemj.org/bj/424/0317/4240317.pdf

5.4.3 Case Kepler

Gruppen bag de stallarseismiske data fra Kepler satelitten har besluttet, at alle skal have mulighed for at arbejde med data med det samme, men at andre end de, som har stået bag observationerne først må publicere deres resultater efter 6 måneder. De, som arbejder med data, er også med i et community, hvor man kan komme med forslag til projekter, anmode om at arbejde med et udvalg af data og submitte artikler til internt "preview" før de sendes til tidsskrifter. Formålet med reglerne er på den ene side at tilskynde til åbenhed og på den anden side beskytte de interesser, som der er i data.

Resultatet er et spind af relationer – mellem personer, datasæt, projektforslag, preprints og artikler. En første version af et site er fungerende, men designet har

helt klart haft fokus på de informationer, som er nødvendige her og nu. Et mål er fremadrettet at etablere relationer, som sikrer, at data også er tilgængelige og brugbare fremover.

Gruppen vil altså sikre fremtidig adgang til og brug af datasættene og er derfor indstillet på at arbejde med metoder, til så automatisk at indfange alt relevant information om sammenhænge og forudsætninger.

5.5 Linkning og citerbarhed af forskningsdata

Der er mange løsninger der bejler om at løse opgaven om at identificere digitale objekter, f.eks. helt generelle URI, URL'er til ellers ukendte ARK-identifikatorer. Disse er alle blevet vurderet i projektet ud fra litteraturstudier.

På nationalt plan arbejdes der i et andet DEFF-projekt på at etablere en national løsning som styres af KB og SB. Der er to mulige bejlere, URN og DOI'er. URN viser sig at være adopteret af de nationale biblioteker og ønskes herfra, mens DOI er mere kendte i forskningsbibliotekerne.

I nærværende projekt er DOI undersøgt som den førende løsning for forskningsdata. DTIC er medlem i DataCite og har derfor været involveret igennem hele skabelsen og har derfor gennemført de her nævnte analyser.

DataCite er en organisation som har til formål at bygge en infrastruktur og best practices, som muliggøre alle de services der kendes fra publiceringer til forskningsdataområde. Det er søgning, citering og forskningsevaluering gennem f.eks. Web of Science eller Scopus.

DataCite har igennem sit korte liv genereret en første praksis på området og har derigennem skabt en fælles de-facto standard. Der er etableret interfaces til online-registrering af DOI'er og deres tilhørende "lokalisering" i form af en URL. Disse kan vedligeholdes gennem samme interface. DataCite har etableret en citationsstandard og en metadata standard, der er holdt på en måde at der kræves så lidt som mulig for registreringen, men samtidig er stor nok til at beskrive specielle forhold og detaljer. Organisationen er nu gået fra at etablere de helt basale services til at udvikle og tilbyde ekstra services, som en katalog over alle DataCite-ressourcer, oversigter over datacentre og meget mere. Disse bliver prioriteret af en international arbejdsgruppe, som nok også skal holde sig til DataCites roadmap.

I nærværende projekt er der anvendt DataCite på en indirekte måde, da der ikke var ressourcer der kunne registreres i DataCite, der kræver "persistens". Der blev derfor taget digitale objekter fra konferencesystemet, der it-arkitektonisk ligner fuldstændig til forskningsdata-installationen. Manuelle forsøg blev aflyst af command-line styrede scripts der tilgår den internationale registreringssystem. For tiden bliver der arbejdet på en automatisering af dette proces. Der er registret 2000 DOI'er på en gang uden problemer. De bliver løbende overvåget for at se om der opstår fejl, da der ikke har erfaringer med sådanne registreringer.

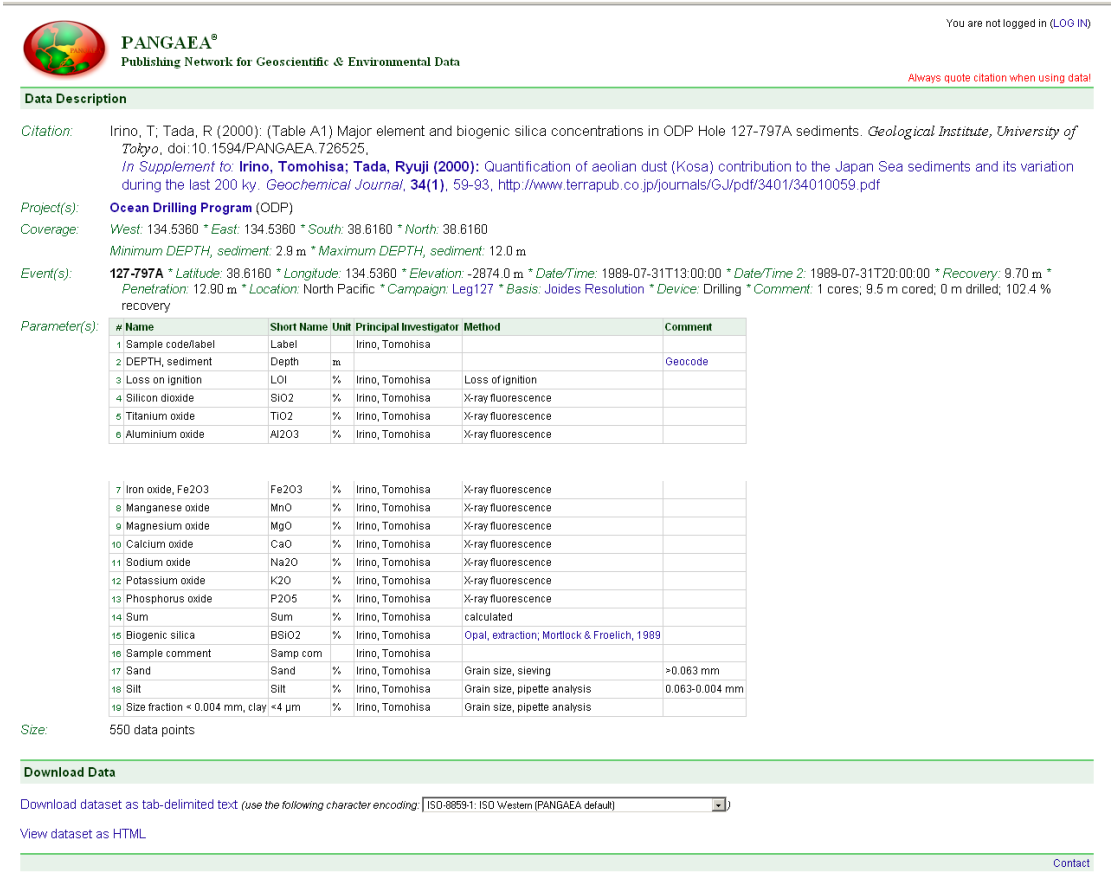
Løsningen er nem at bruge, hvis man kan sende sine data i DataCites metadataformat. Man kan nemt registrere manuelt, mens automatiske

registreringer skal indbygges i softwaren for den aktuelle case eller omkring den. F.eks. vil man skulle udvide DataVerse for at kunne få automatisk integreret med DOI-identifikatorer.

Citering af publiceringer er veletableret og har dannet grundlag for tilsvarende løsninger på forskningsdataområdet. Speciel TIB, Hannover har spillet en afgørende rolle og var initiativbærer for DataCite og dermed anvendelse af DOI'er til identifikation og citering af forskningsresultater.

DataCite har udviklet en metadata-kærne som består af 6 obligatoriske felter samt 3 ekstra felter til citering. Dermed kan man fra publiceringen referere til forskningsdata fuldstændig som forskerne gør det for at pege på publiceringer. Da DOI'er kan "resolves" kan referencen gøres "klikbar" i de digitale medier for begge retninger – fra publikationen til forskningsdata og omvendt. Et eksempel fra Jan Brase, DataCite, gengives her som et eksempel:

Figuren viser en artikel fra ScienceDirect, hvor forlaget skanner deres artikler for DOI'er som de kan genkende, her DataCite-DOI'er, mere præcis DOI'er som er tildelt for projektet PANGEA i Tyskland. På højre nederste del af figuren ses, at der refereres til "Supplementary Data" gennem DOI-resolveren, hvor DOI'en har følgende streng http://dx.doi.org/10.1594/PANGAEA.724325. En sådan DOI kan pege på selve dataene, men i næværenden eksempel, samt de fleste, så peger man i stedet på de beskrivende metadata – se næste figur:



PANGAEA®
Publishing Network for Geoscientific & Environmental Data

You are not logged in (LOG IN)

Always quote citation when using data!

Data Description

Citation: Irino, T; Tada, R (2000): (Table A1) Major element and biogenic silica concentrations in ODP Hole 127-797A sediments. *Geological Institute, University of Tokyo*, doi:10.1594/PANGAEA.726525,
In Supplement to: **Irino, Tomohisa; Tada, Ryuji (2000):** Quantification of aeolian dust (Kosa) contribution to the Japan Sea sediments and its variation during the last 200 ky. *Geochemical Journal*, **34(1)**, 59-93, http://www.terrapub.co.jp/journals/GJ/pdf/3401/34010059.pdf

Project(s): **Ocean Drilling Program** (ODP)

Coverage: *West:* 134.5360 * *East:* 134.5360 * *South:* 38.6160 * *North:* 38.6160
Minimum DEPTH, sediment: 2.9 m * *Maximum DEPTH, sediment:* 12.0 m

Event(s): **127-797A** * *Latitude:* 38.6160 * *Longitude:* 134.5360 * *Elevation:* -2874.0 m * *Date/Time:* 1989-07-31T13:00:00 * *Date/Time 2:* 1989-07-31T20:00:00 * *Recovery:* 9.70 m * *Penetration:* 12.90 m * *Location:* North Pacific * *Campaign:* Leg127 * *Basis:* Joides Resolution * *Device:* Drilling * *Comment:* 1 cores; 9.5 m cored; 0 m drilled; 102.4 % recovery

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	Sample code/label	Label		Irino, Tomohisa		
2	DEPTH, sediment	Depth	m			Geocode
3	Loss on ignition	LOI	%	Irino, Tomohisa	Loss of ignition	
4	Silicon dioxide	SiO2	%	Irino, Tomohisa	X-ray fluorescence	
5	Titanium oxide	TiO2	%	Irino, Tomohisa	X-ray fluorescence	
6	Aluminium oxide	Al2O3	%	Irino, Tomohisa	X-ray fluorescence	
7	Iron oxide, Fe2O3	Fe2O3	%	Irino, Tomohisa	X-ray fluorescence	
8	Manganese oxide	MnO	%	Irino, Tomohisa	X-ray fluorescence	
9	Magnesium oxide	MgO	%	Irino, Tomohisa	X-ray fluorescence	
10	Calcium oxide	CaO	%	Irino, Tomohisa	X-ray fluorescence	
11	Sodium oxide	Na2O	%	Irino, Tomohisa	X-ray fluorescence	
12	Potassium oxide	K2O	%	Irino, Tomohisa	X-ray fluorescence	
13	Phosphorus oxide	P2O5	%	Irino, Tomohisa	X-ray fluorescence	
14	Sum	Sum	%	Irino, Tomohisa	calculated	
15	Biogenic silica	BSiO2	%	Irino, Tomohisa	Opal, extraction; Mortlock & Froelich, 1989	
16	Sample comment	Samp com		Irino, Tomohisa		
17	Sand	Sand	%	Irino, Tomohisa	Grain size, sieving	>0.063 mm
18	Silt	Silt	%	Irino, Tomohisa	Grain size, pipette analysis	0.063-0.004 mm
19	Size fraction < 0.004 mm, clay	<4 µm	%	Irino, Tomohisa	Grain size, pipette analysis	

Size: 550 data points

Download Data

Download dataset as tab-delimited text *(use the following character encoding:* ISO-8859-1: ISO Western (PANGAEA default)*)*

View dataset as HTML

Contact

Til citeringen vil man kunne anvende DataCites format som er givet på følgende måde:

> Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. Geological Institute, University of Tokyo. doi:10.1594/PANGAEA.726855. http://dx.doi.org/10.1594/PANGAEA.726855

Som man kan se, ligner det fuldstændig den måde man citerer publikationer på. Første streng der starter med doi:10. er selve doi, mens tilsvarende URL-version er tilføjet for at gør doi'en "resolvable" (klickbar).

Bemærk at DOI'er (og andre persistente identifikatorer) kan tildeles ikke kun data og publiktioner, men deres dele som f.eks. figurer, tabeller osv. Dermed kan man meget præcis pege ind på en given forskningsresultat.

Lignende danske eksempler er etableret, men publikationerne er ikke udgivet endnu for at vise i nærværende dokument.

5.6 Metadata til beskrivelse af forskningsdata

Der findes mange metadataformater, også for forskningsdata. DDI anses som den mest fremtrædende standard på området, men den er meget krævende og mangler åbne og nemt anvendelige software til inddatering af informationerne. Der arbejdes netop nu på at udvikle en lang række åbne programmer til brug for skabelsen af datadokumentation jf. DDI standarden. Der er udviklet et enkelt

kommercielt produkt fra Colectica (hjemmeside: http://www.colectica.com/), der kan anvendes til at generere DDI-dokumentationen.

DataCite er kommet op med et alternativ, der ikke kræver meget, men kan medtage ret omfattende beskrivelser, hvis dette ønskes. Formatet er defineret til at kunne etablere citeringer af forskningsdata og –resultater.

Det australske ANDS har udviklet RIS-CS som er en ISO-standard, som er udviklet til andre formål og kan anvendes til forskningsdata. Denne standard er overskuelig og kan anvendes til flere formål.

Dublin Core har samme begrænsninger på forskningsdataområdet, som den har til at beskrive publiceringer. Det er nævnt at DataCite's metadata ligner DC meget.

Det kræver mange flere forsøg for at vurdere disse metadataformater til beskrivelse af forskningsdata. Man kan dog sige, at DataCite's metadataformat ikke kan beskrive de enkelte målingernes egenskaber og derfor skal man videre til RIS-CS, hvis man vil beskrive dataene i detaljer. Om DDI kan anvendes, afhænger af, om der skabes værktøjer til at hjælpe til at etablere disse komplicerede metadata.

DTIC har i sin Fedora-installation anvendt DataCite for registrering af DOI'er, RIS-CS for at afgøre RIS-CS formatet, DC, men ikke DDI der bliver bredt anvendt på DDA. Konklusionen er, at der behøves flere metadatabeskrivelser for hver sit formål. Ingen, ud over DDI-formatet, beskriver detaljer omkring dataenes format, hvilket man skal beskrive på andet vis.

5.7 Synliggørelse af forskningsdata

Før der gennemgås de resultater der er evalueret i projektet, er det relevant at henvise til den wiki som projektet har frembragt og hvor man kan finde mange andre eksempler på synliggørelse af forskningsdata. Eksemplerne er så omfattende og var inspiration til implementering af Fedora-demonstratoren på DTIC.

Projektet har undersøgt forskningsdatabaserne (Institutional Repositories) for en mulig løsning for håndtering af forskningsdata. Konkret blev DTU's Orbit og PURE-softwaren der anvendes på alle andre universiteter, undersøgt. Resultatet er klart, at disse ikke, på nuværende stadie, kan håndtere opgaven – de er bygget til andre formål og har alt for simplificerede modeller for de digitale objekter der håndteres.

Dedikeret software for forskningsdata er nødvendig, da forskningsdata ikke ligner publikationer og andre forskningsresultater. Hvis forskningsdata skal understøttes med services og beskrivelser på de niveauer af brug der forekommer i hverdagen, så kræves det disse specialsoftware og IR kan ikke anvendes til formålet. Nedenfor gennemgås de mest omtalte løsninger til håndtering af forskningsdata.

5.7.1 DataVerse

Dataverse er et Open Source program, der er udviklet på Harvard. I Dataverse kan man lægge data, der kommer fra mange forskellige formater. Formålet med

Dataverse er at bevare data, dele dem med andre og at kunne (gen)finde data. En institution kan være vært for andre institutioner, der får deres eget område på den samme centrale installation med deres egne logoer etc. Dette kan være en stor fordel, da installationen af Dataverse kan være lidt vanskelig.

Der er et brugbart interface til at beskrive baggrundsinformationen til en undersøgelse. Den giver mulighed for at beskrive en undersøgelse på samme måde, som det allerede finder sted ved de samfundsvidenskabelige dataarkiver – men den kan også bruges til andre typer forskningsdata, idet der er tale om generel baggrundsinformation om data. Der er mulighed for at lægge alle former for dokumenter og data op i Dataverse og dermed dele dem med andre. Ved bestemte formater som SPSS og Stata tilbyder Dataverse muligheder for online analyser. I andre tilfælde kan man downloade data og foretage analyserne på sin egen computer. Man bestemmer selv adgangen til de enkelte dokumenter. Således kan der være fri adgang til datadokumentationen, men ikke til selve datafilerne.

Metadata i Dataverse Network følger DDI version 2 (www.ddialliance.org), der kan eksporteres i XML format. Kompatibelt med simple Dublin Core krav (dublincore.org). Import kompatibelt med Content Standard for Digital Geospatial Metadata (CSDGM), Vers. 2 (FGDC-STD-001-1998) (FGDC) (www.fgdc.gov/metadata/csdgm). For yderligere information se http://thedata.org/sites/projects.iq.harvard.edu/files/CatalogingFields11Apr08.pdf.

Installationen af programmet er sket hhv. på KUBIS der evaluerer programmet på nuværende tidspunkt. Første erfaringer viser:

I 2009 gennemførte KUBIS en mindre, kvalitativ undersøgelse om samfundsvidenskabernes indsamling, brug og udveksling af primærdata. Rapporten (www.hprints.org/hprints-00451000), der udkom i januar 2010, konkluderer blandt andet, at der er et udtalt behov hos de adspurgte forskere for backupprocedurer og sikker arkivering af forskningsdata, samt behov for at forskerne kan dele data med andre forskere, studerende eller andre personer, og at de selv kan styre, hvem der har adgang.
Derfor undersøges der ved dette projekt om softwaren "Dataverse" vil kunne opfylde dette behov for data backup og deling, med henblik på senere høstning til langtidsbevaring.
Et Dataverse er et online dataarkiv, hvori forskerne kan uploade data via et webinterface og hvor de kan vælge at lægge data åbent ud til offentligheden, holde dem private, eller dele dem med udvalgte kolleger og studerende.

DDA har ikke selv installeret Dataverse, men benytter p.t. den installation Sveriges Nationale Datatjänst (SND) har. Jeg har fra både SND og fra det amerikanske ICPSR er det forstået, at installationen af DataVerse kræver en del korrespondance med udviklerne for at få det til at fungere. I SND's Dataverse Network har Dansk Data Arkiv sit eget Dataverse med alle rettigheder og muligheder for at tilpasse det efter behov. På denne adresse er der adgang til DDA's dataverse:
http://130.241.110.43/dvn/dv/DDA;jsessionid=6e8acfe5a2e2578032ef46c16d3c

Der er lagt to datamaterialer og et dokument op. Denne del af processen er forløbet rimelig uproblematisk. Der er et brugbart interface, til at beskrive baggrundsinformationen til en undersøgelse. Den giver mulighed for at beskrive en undersøgelse på samme måde, som det allerede finder sted ved de samfundsvidenskabelige dataarkiver – men den kan sandsynligvis også bruges til andre typer forskningsdata. Der er mulighed for at lægge de fleste former for dokumenter op i DataVerse og dermed dele dem med andre. Man bestemmer selv adgangen til de enkelte dokumenter.

Som med alle programmer er der en del funktioner man skal lære og der er nogle begrænsninger, men det overordnede er, at det er muligt at lægge de fleste typer – måske alle – op og vise dem til andre. Download-funktionen er mere værdifuld end muligheden for at lave online analyser. DDA's erfaringer med den aktuelle udgave hos SND viser, at installationen af programmet er en af de største udfordringer.

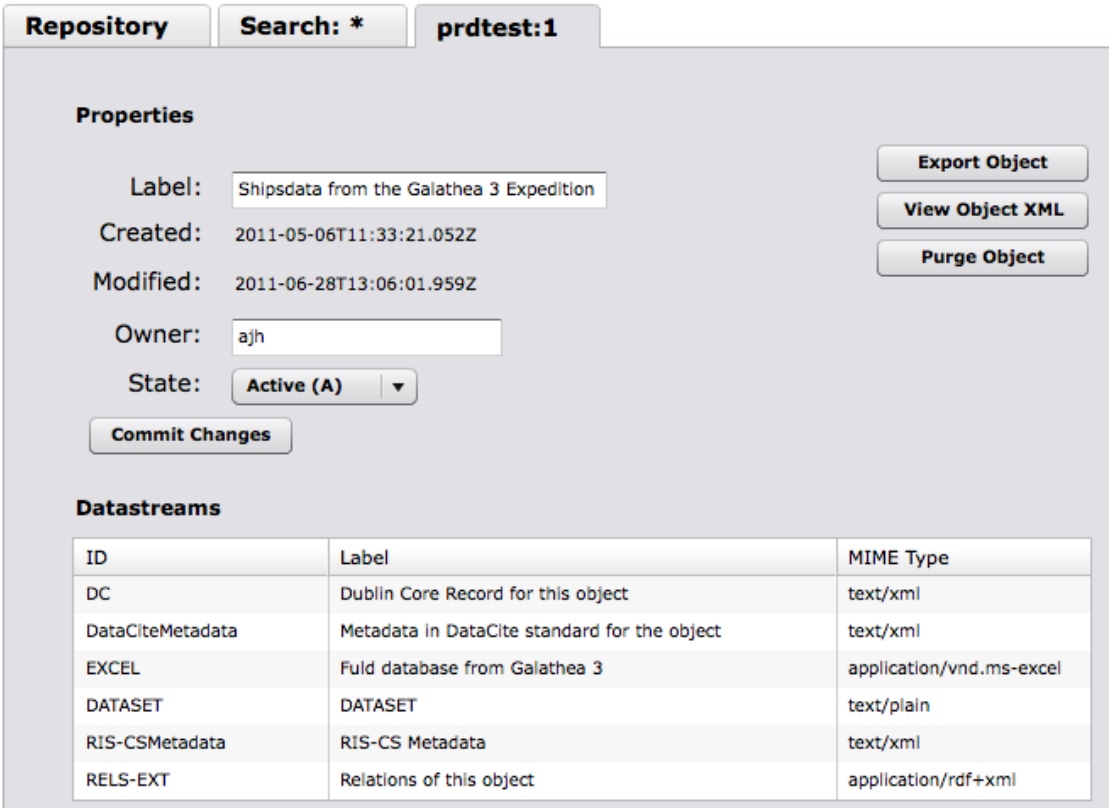
Internationale erfaringer med Dataverse er positive, lokale erfaringer er under udarbejdelse. Dataverse er et meget ordnet system med god mulighed for strukturerede metadata. Denne mulighed for metadata gør dog også at det fra brugerens side kunne opfattes som et lidt tungt system, og der kunne muligvis stilles ønske om et system, der f.eks. faciliterer drag-and-drop overførsler, men det vil næppe være et system, der i samme grad som Dataverse vil understøtte metadata.

Der er et potentielt stort problem i at Dataverse softwaren ikke understøtter krypteret up- og download, hvilket betyder at systemet måske er mindre egnet til personfølsomme data. Dette kan betyde at de forskere der måske mest har brug for back-up hjælp fordi de ikke kan bruge kommercielle online services såsom skydrive eller dropbox pga person- eller af anden årsag følsomme data, heller ikke vil bruge Dataverse, men det vil vise sig ved kommende interviews.

5.7.2 Fedora-Commons til håndtering af forskningsdata

DTIC har videreudviklet på open source platformen Fedora-Commons. Løsningen kan bruges til at indlægge data, metadatabeskrive dataene, sætte relationer mellem disse og andre objekter, samt levere dataene gennem et web service interface. Det hele styres gennem et ret simpelt web-administrations-grænseflade eller på "command-line" af teknisk personale.

Her et billede af administrationsgrænsefladen:

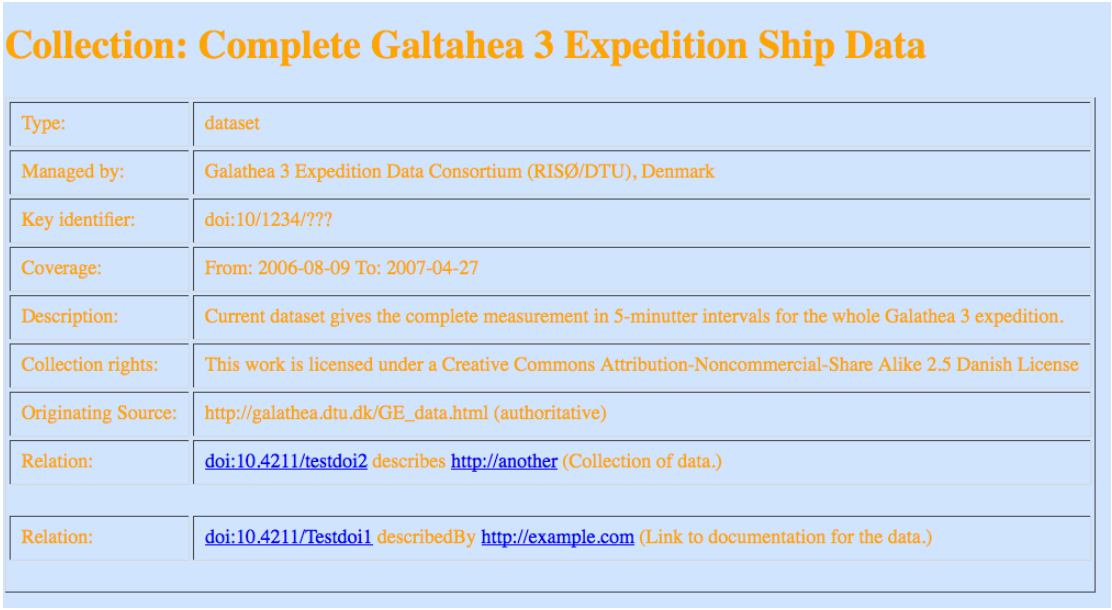


Figuren viser en repræsentation af de dele som forskningsdataobjektet består af. Det kan fornemmes af navngivningen, at der er en Dublin Core beskrivelse (metadata) som bruges af Fedora selv og kan udnyttes til andre formål som OAI-høstning, en beskrivelse i DataCite-format (til registering af DOI'er), datasettet i Excel-format og som tal, en metadatabeskrivelse i RIS-CS-format, samt et stykke information der beskriver relationerne mellem disse dele af objektet og også beskrivelser der ligger udenfor objektet, herunder beskrivelsen af dataformatet som sådan (hvad er målt, hvor præcis osv).

Ønskes f.eks. rådataene ladet ned på sin computer, klikkes på den givne link og man får det direkte ind i Excel da mime-typen er angivet. Tilsvarende kan man fra ethvert andet interface eller browser kalde den enkelte del eller hele oversigten gennem en link som f.eks. http://severnavn.dk/fedora/objects/prd1:1000000002/datastreams/DATASET/content

Som man kan se, er sådanne links ret komplicerede og en persistent identifikator vil ledte ens arbejde i forhold til linkning.

En første ide af en visualisering af metadata for en given forskningsdataobjekt kan ses gennem en prototype af et frontend til formålet, som danner grundlag for det videre arbejde på DTIC:



Collection: Complete Galtahea 3 Expedition Ship Data

Type:	dataset
Managed by:	Galathea 3 Expedition Data Consortium (RISØ/DTU), Denmark
Key identifier:	doi:10/1234/???
Coverage:	From: 2006-08-09 To: 2007-04-27
Description:	Current dataset gives the complete measurement in 5-minutter intervals for the whole Galathea 3 expedition.
Collection rights:	This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 2.5 Danish License
Originating Source:	http://galathea.dtu.dk/GE_data.html (authoritative)
Relation:	doi:10.4211/testdoi2 describes http://another (Collection of data.)
Relation:	doi:10.4211/Testdoi1 describedBy http://example.com (Link to documentation for the data.)

Samme fremgangsmåde er anvendt af ANDS discovery interface der udstiller på følgende måde: http://services.ands.org.au/home/orca/rda/view.php?key=102.100.100%2F4522.

Fordelen med løsningen er, at man kan modellere utrolig komplicerede forhold i dataene og fremvisning af denne, hvor f.eks. Dataverse har et veldefinerede og derigennem begrænset bud på, hvordan data i hele taget skal håndteres. Ulempen er, at brugergrænsefladerne til f.eks. inddatering af metadata skal gøres i andre værktøjer der er velegnet til formålet og herefter overføres til Fedora-løsningen.

Da IT-arkitetkturen i Fedora-systemet er den samme som den der anvendes til konferencesystemet på DTU, kan man referere til dette for at se det "in action" – et datakollektion kan sammenlignes med en konference, et datasæt kan sammenlignes med en præsentation/paper osv. Fedora udstiller disse digitale objekter på en relevant måde i en beskrivende side hvorfra man klikke sig til selve data-objekterne. Man ville kunne finde gennem metadataene til data-centre, forskere og deres andre arbejder osv. En stor forskel vil være i forhold til beskrivelser, hvor man for publikationer og præsentationer kender til at afkode, hvad objekterne indeholder, er dette ikke tilfældet for forskningsdata – her er der behov for "faglig metadata" der bl.a. beskriver hvad man måler, i hvilke præcision de er gengivet, hvilke instrumenter der har frembragt dataene og under hvilke betingelser (f.eks. kalibrering). For disse data skal der ekstra præsentationer til systemet.

Under bedst practice er andre systemer og mange eksempler for fremvisning af data foreslået – se wikien som projektet har frembragt.

5.7.3 Andre løsninger

På basis af Fedora-Commons er der udviklet eSciDoc som udvikles til de processer der foregår i forskningen, herunder et "institutional repository", en system til samlinger og relevant her, et arkiv til håndtering af forskningsdata.

Hydra er et andet projekt under Fedora-Commons som sandsynligvis vil kunne bidrage til håndtering af forskningsdata og forskningsprocesser.

Det nyeste bud bliver i den kommende tid udviklet med støtte fra det amerikanske NSF, LabArchives.

Flere nævnes under Best Practice.

5.8 Forskerkontakten er stadig udfordringen

Selvom bibliotekerne anerkendes som troværdige partner for forskerne, så viser det sig i projektet, at samarbejdet omkring forskningsdata ikke gav de nødvendige resultater. Samarbejdet omkring Kepler-data viser sig at overgå tidsrammen for nærværende projekt. Samme for Tor-projektet. På DTIC er der gennemført flere samarbejder der strandede på flere forskellige måder. At dataene efter arkiveringen ville holdes hemmeligt af enkelte forskere, at dataene ikke kunne tilvejebringes af flere grunde, at tilvejebringning af data var for tidskrævende for de involverede forskere og meget mere. Alt i alt kan man opleve viljen fra forskernes side, men tidspresset på dem forhindrer, at dataene sikres på forsvarlig vis.

6 Best Practice – Internationale erfaringer

Formålet med nærværende best practice rapport for forskningsdata og open access er at skabe et overblik og en erfaringsopsamling af udvalgte cases, som kan tjene som guidelines i det fremadrettede arbejde med forskningsdata og open access i Danmark. De udvalgte cases er med, fordi der er vurderet, at de kan have en praktisk værdi som inspirationskilde for bibliotekssektoren og samarbejdspartnere der ønsker at initiere projekter på dette område. Der er ikke forsøgt at lave en udtømmende dækning af området, men udelukkende søgt efter cases som vurderes, kan bruges som inspiration og afsæt i forhold til en udviklingsproces, som DEFF's interessenter ønsker at involvere sig i for forskningssektoren som helhed.

6.1 Metode til indsamling af best practice cases og relevant litteratur

Oversigten består dels af korte beskrivelser af relevante internationale og nationale projekter med forskningsdata, dels en linksamling og en samling af relevante rapporter, strategipapirer, politiknotater mv. Det er tanken, at man via denne rapport på wikien kan danne sig et overblik over feltet via de cases som er fokuseret på. De enkelte cases er udvalgt fordi de er interessante og relevante som spydspids cases og qua deres forskellige løsningsmodeller. Disse cases belyser forskellige muligheder og udfordringer på forskningsdataområdet samt hvilke indsatsområder der er relevante for forskellige aktører som forskere, biblioteker, nationale organer og myndigheder, systemudbydere mv.

Ved hjælp af netværk og kilder på nettet har vi forsøgt at afdække de væsentlige internationale erfaringer udover det materiale som allerede ligger i dette projekts kildegrundlag. Det har resulteret i følgende oversigt over:

- internationale cases
- systemer til håndtering af datasæt

- rapporter og strategipapirer
- data-sikkerhed

6.2 Internationale cases

6.2.1 Australian National Data Service (ANDS).

ANDS er både et infrastruktur- og cultural change project for forskningsdataområdet på nationalt plan i Australien. De australske erfaringer er kendetegnet ved et omfattende forarbejde omkring mål ,vision og strategi til en national forskningsdatabank. Data sharing forstås i det australske projekt som C*reate, Store, Identify, Describe, Register, Discover, Access and Exploit.* Disse *"data sharing Verbs"* tegner behov og krav til meningsfulde dataworkflows og systemarkitektur. Data sharing Verbs er fundamententet for den australske best practice. Nedenfor linkes til forundersøgelser og artikler, der beskriver den Australske case via Andrew Treloars hjemmeside. Andrew Treloars er lead agent for Australien National Data Service (ANDS). Artiklerne beskriver detaljeret baggrund og indsats omkring udrulningen af en national dataservice for opsamling og formidling af forskningsdata i Australien. for-undersøgelser og baggrund ANDS (4 artikler i boks midt på siden)

Nærværende projekt kan drage stor nytte af de australske erfaringer på flere niveauer. Det er interessant at opnå yderligere indsigt i data workflow i ANDS systemet. Samt at få viden om ejerskab og konkret brug af ANDS i de faglige miljøer og hos de enkelte forskere. Herunder viden om hvor integreret brugen af ANDS er i de akademiske miljøer. Parallelt til den danske indsats på forskningsdataområdet er det helt essentielt, at projektet funderes i de videnskabelige miljøers forskelligartede behov for opbevaring og formidling. De australske erfaringer, best practice og ideer til en handlingsplan for Danmark er beskrevet af Alfred Heller og Lars Nondal i **Revy 1-34** (side 16-18). Artiklen er et nyttigt bidrag til arbejdet med forskningsdata i dansk kontekst.

Nøglepersoner og samarbejdsaftaler ANDS:

Andrew Treloar, Deputy Director ANDS med hvilke der blev afholdt møder på turen i Australien (uge 41- 2010). En anden nøgleperson er Sam Searle, Monash University. Monash University er lead institution for ANDS. Sam Searle har erfaringer omkring institutional engagement, workflow og Know How omkring forskningsdata i Australien. Den ovenfornævnte artikel i Revy er en del af afrapporteringen fra projektdeltageres tur til Australien.

6.2.2 DANS - et hollandsk initiativ

Data Archiving & Networked Services (DANS) http://www.dans.knaw.nl/en opbevarer og formidler datasæt indenfor den humanistiske og samfundsvidenskabelige forskning i Holland. DANS har eksisteret siden 2005 og et af kerneområderne er support, koordinering og udvikling af arkivsystemet EASY til opbevaring af datasæt. DANS yder adgang til datasætfiler på nationalt niveau, såvel som på internationalt niveau. De hollandske erfaringer beskrives på DANS's hjemmeside under Publikationer. Her linkes til rapporter, baggrundsmateriale, manualer mv., som kan bidrage med vigtig information og erfaringer i forhold til den danske indsats.

Nøglepersoner Holland
Laurents Sesink er informationsvidensskabelig forsker og ansat på DANS. Han arbejder med adgang og formidling af digitale videnskabelige data. Hans fokus er på hvordan teoretiske løsninger kan implementeres praktisk i R&D projekter i DANS.

6.2.3 Knowledge Exchange

Knowledge Exchange samarbejdet understøtter brug og udvikling af Information- og kommunikationteknologiens infrastruktur for forskning og højere uddannelse. Knowledge Exchange samarbejdet består af følgende aktører:

- Denmark's Electronic Research Library (DEFF) i Danmark
- German Research Foundation (DFG) i Tyskland
- Joint Information Systems Committee (JISC) i England
- SURFfoundation i Holland

Knowledge Exchange samarbejdet arbejder målrettet med projekter indenfor Open Access, Licensing, Repositories, Research Data og Virtual Research Environments. Under Research data delen af Knowledge Exchange pågår der aktiviteter af interesse for forskningsdata, herunder workshoppen *Main drivers for successful re-use of research data* holdt i september 2009. Workshoppens resultater er dokumenteret i KE_Workshop_Berlin_Preliminary_Report_2010-11-12_KR.pdf Da workshoppen var centreret om behov ud fra forskeres optik er rapporten et bidrag i den indledende fase af projekter om forskningsdata og open access. Indenfor KE er der ligeledes nedsat en Primary Research Data Working Group med eksperter fra de 4 partner organisationer.

6.2.4 Joint Information Systems Commitee (JISC)

JISC er den engelske partner i Knowledge Exchange samarbejdet. JISC har som partner i KE søsat projekter omkring forskningsdata, og har endvidere i eget regi udviklet en lang række initiativer på forskningsdataområdet. I det følgende trækkes de vigtigste aktiviteter frem med relevans for denne kontekst. Et af JISC vigtigste indsatsområder er at bane vejen for god management og formidling af forskningsdata til fordel for forskning og akademisk uddannelse i England. JISC har således udvikllet *The Managing Research Data Programme* JISCMRD som adresserer en lang række fokusområder på forskningsdata-området, herunder:

- Piloting essential research data management infrastructures within institutions and for distributed research groups
- Improving practice in research data management planning
- Developing tools to help institutions plan their research data management practice
- Encouraging the publication of research data and demonstrating the benefits of improved methods for citing, linking and integrating research data
- and, stimulating the acquisition of appropriate skills, among academics and research support staff in Universities

For at opnå resultater indenfor områderne har JISC udarbejdet fem program strenge som indeholder projekter centreret om forskningsdata, open access og kompetenceudvikling. Rapporterne under de fem programmer er således relevante for udviklingsprojekter omkring forskningsdata i nærværende kontekst.

1. Research Data Management Infrastructure (RDMI) Projects
2. Research Data Management Planning (RDMP) Projects
3. Support and Tools Projects
4. Citing, Linking, Integrating and Publishing Research Data (CLIP) Projects
5. Research Data Management Training Materials Projects

6.2.5 Digital Curation Centre

JISC finansierer Digital Curation Centre som er et engelsk center for organisering af digitale forskningsdata. Centeret tilbyder hjælp til effektiv data management gennem alle faser af forskningskredsløbet for at sikre optimal produktivitet og genbrug af forskningsresultater. På Digital Curation Centre's hjemmeside, er der hjælp at hente i form af litteratur og projekter (under fanebladene *Ressourcer* og *Projects*) ligesom der foreligger praktiske gudelines til optimering af daglige arbejdsgange med forskningsdata How to guides DDC har ligeledes udviklet relevant trænings- og uddannelsesmateriale for aktører indenfor forskningsdatafeltet - både for forskere, bibliotekarer og support-personale (under fanebladet *Training*). Nedenstående FAQ kan guide interesserede for hjælp til specifikke services, emner og målgrupper:

- Data Creators - researchers and principal investigators
- Data Managers and research liaison staff embedded at research group or department level
- Data Librarians, Computing & IT Service Managers
- Data Scientists and Library or Informatics Researchers
- Senior Research Managers
- Funding Bodies, National Data Centres, Learned Societies & Professional Bodies
- Consultants

6.3 Systemer til håndtering af datasæt

6.3.1 The Dataverse Network

The Dataverse Network er en Open Source Application til at publicere, citere, analysere og formidle forskningsdata. Det overordnede formål er lagring og genbrug af data med henblik på at facilitere innovation og videndeling. Et Dataverse netværk er vært for mange Dataverses. Hver Dataverse indeholder

studier eller samlinger af studier som indeholder katalogiseringsinformation om data, samt aktuelle datafiler og supplerende filer. Systemet tilbyder den enkelte forsker eller forskningsinstitution en mulighed for at oploade data i IQSS Dataverse Networksom er hostet at IQSS, Harvard Universitet. Det er frit tilgængeligt for alle dataskabere og distributører af data i verden. Den enkelte forsker kan her udarbejde sit eget Dataverse og inddatere sine forskningsdatasæt og supplerende data som dokumenter, software mv. Det er også muligt for den enkelte institution at hoste sit eget Dataverse Network. Denne installationsmulighed henvender sig kun til de organisationer som vælger at være vært for egen Dataverse Network Software.

Datamanagement på Massachusetts Institute of Technology Libraries (MITLibraries).

På MITLibraries er der udarbejdet en step-by-step guideline til håndtering af forskningsdata-cyklussen fra fødsel til distribution i systemer. Den definerer hvad data er og hvilke minimumskrav, der må stilles for beskrivelse af data. Sitet indeholder en Data Planning Check Liste der punkt for punkt lister de forhold, som der skal tages højde for i planlægningen af arbejdet med forskningsdata.

Det er en 17 punkts liste, der giver konkrete anvisninger i forhold til planlægning og arbejde med lagring, genfinding, relationsopbygning mellem dokumenter, samt muligheder for genbrug og formidling af datasæt.

MITLibraries er et eksempel på et forskningsbibliotek, der har udarbejdet en service for deres brugere, og som har guidelines der er frit tilgængelige til inspiration for andre der vil arbejde i samme retning. For forskere og studerende tilknyttet MIT er der forskellige muligheder for at lagre, genfinde og dele data indenfor forskellige domæner.

6.4 Rapporter, strategipapirer m.m.

Efter afslutningen af projektet blev der udgivet en foreløbig rapport om de opgaver der ligger for på dataområdet, udgivet af en international samarbejde – Dette arbejde ses som central i arbejdsfeltet og refereres derfor i starten af dette afsnit:

”GRDI2020 –Towards a 10-year Vision for Global Research Data Infrastructures”, hvor der skrives: ”We envision that, in the immediate future, Interoperable Science Ecosystems, composed of Digital Data Libraries, Digital Data Archives, and Digital Research Libraries, will be established and Global Scientific Data Infrastructures will act as the enablers of these interoperable science ecosystems.

6.4.1 SURF rapporten

Med rapporten **"What Researchers want"** har SURF foretaget et litteraturstudie af forskeres behov i forhold til forskningsdata. Der er inddraget rapporter der dækker forskellige perspektiver i feltet omkring forskningsdata. Disse perspektiver dækker organisering, infrastruktur, forskeres behov, open access og policy. Udgangspunktet for rapporten er, at det er essentielt at afdække forskernes behov, før der udvikles systemer eller processer indenfor området. En konklusion er, at der skal tages hensyn til forskernes ønske om egen kontrol

med datasæt i ft. adgang, og samtidig hvilke områder der kan være gavnlige at fokusere på som støttefunktioner for forskerne.

Nærværende rapport konkluderer: "...that researchers can, indeed, benefit from support services in managing their digital data, but that these services must meet a number of requirements if they are to be successful:

- Tools and services must be in tune with researchers' workflows, which are often discipline- specific (and sometimes even project-specific).
- Researchers resist top-down and/or mandatory schemes.
- Researchers favour a "cafeteria" model in which they can pick and choose from a set of services.
- Tools and services must be easy to use.
- Researchers must be in control of what happens to their data, who has access to it, and under which conditions. Consequently, they want to be sure that whoever is dealing with their data (data centre, library, etc.) will respect their interests.
- Researchers expect tools and services to support their day-to-day work within the research project, and long-term/public requirements must be subordinate to that interest.
- The benefits of the support must be clearly visible – not in three years' time, but now.
- Support must be local, hands-on, and available when needed.

Rapporten giver et overblik over litteratur på feltet, og dermed inspirerer til videre studier. Endvidere fokuserer rapporten på en grundig afdækning af forskernes behov, hvilket er afgørende viden for det videre udviklingsarbejde omkring forskningsdata og OA. En "Must read"-rapport for projektmagere indenfor forskningsdataområdet.

Kilde: SURFfoundation, www.surf.nl , **"What Researchers want"**

6.4.2 CARL rapporten

CARL rapporten Addressing the Research Data Gap - A Review of Novel Services for Libraries.pdf er et litteratur- og casestudie udarbejdet af Canadian Association of Research Libraries (CARL) som gennemgår udvalgte forskningsdataservices internationalt og opdeler dem i 5 overordnede kategorier. De 5 temaer i rapporten er:

- awareness and advocacy
- support and training
- access and discovery
- archiving and preservation

- virtual research environments

Hvert tema indeholder en generel beskrivelse samt en række praktiske eksempler fra forskningsdataprojekter som illustrerer mangfoldigheden i de forskellige projekter og deres udgangspunkt og kontekst.

Kilde: http://www.carl-abrc.ca/about/about-e.html

6.4.3 EU Kommissionen

Rapporten Riding the wave - How Europe can gain from the rising tide of scientific data - Final report of the High Level Expert Group on Scientific Data - October 20 er udarbejdet af 12 europæiske eksperter indenfor forskellige videnskabelige domæner. Rapporten beskriver long term scenarier og muligheder og udfordringer vedrørende adgang til videnskabelige data, lagring og opbevaring af data. Samt strategi og nødvendige tiltag for at realisere forskningsdatavisionen. The High-Level Group on Scientific Data og rapporter herfra tilgås via **Cordis website** som er den Europæiske Unions gateway til forskning og udvikling. Cordis indeholder andre rapporter med erfaringer indenfor E-science og e-infrastruktur aktiviteter.

6.4.4 NSF National Science Foundation

National Science Foundation opfordrer til open access på forskningsdata-området.

"Science is becoming data-intensive and collaborative," noted Ed Seidel, acting assistant director for NSF's Mathematical and Physical Sciences directorate. "Researchers from numerous disciplines need to work together to attack complex problems; openly sharing data will pave the way for researchers to communicate and collaborate more effectively."

Hentet 23.2.2011 på nsf.gov - National Science Foundation (NSF) News - Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans - US National Science Foundation (NSF)

NSFs beslutning hilses velkommen af det amerikanske forskningsbiblioteksmiljø, som ser en væsentlig rolle for forskningsbibliotekerne i forvaltning og formidling af forskningsdata. Således peger ARL (Association of Research Libraries) blandt andet på følgende udvalgte punkter fra http://www.arl.org/rtl/eresearch/escien/nsf/leadershiproles.shtml:

· Kontaktbibliotekarer er fortrolige med deres fakulteters behov for forskningsdata

· NSFs beslutning er godt nyt for udviklingen af innovative publikationsmodeller

· Bibliotekerne leverer i stadig højere grad rådgivning vedrørende forskningsdata

· Forskningsdataområdet er et vigtigt element i næste generations biblioteksservice

· Åben forskningsdataforvaltning og –formidling er et paradigmeskift for bibliotekerne

6.4.5 Data-sikkerhed

Sikring af forskningsdata er endnu en side af sagen, som diskuteres blandt parterne på forskningsdataområdet. At kontrollere og sikre adgangen til forskningsdata er i stigende grad et krav fra såvel bevillingsyderne som dataejerne.

6.4.5.1 DS484

Danske universiteter har lige som alle statslige organisationer siden 2007 skullet implementere DS484, som grundlag for sikkerhedspolitik i den konkrete relevante kontekst. Universiteterne har formuleret sikkerhedspolitikker, som er blevet implementeret, for at sikre at adgang til data ikke kompromitteres. Det vil sige, at der på organisationsniveau allerede er implementeret datasikkerhedspolitikker. Datasikkerhedspolitikker og organisation bag bør inddrages i forbindelse med projekter, der retter sig mod forskningsdata i open access perspektivet.

7 Resultater - Opsummering

Projektet har leveret følgende resultater:

1. Input til DEFF-styregruppe, strategi og handlingsplan der bl.a. resulterede i konkrete tilpasninger af visionen, handlingsplanen, samt blev opstart til to konkrete projektansøgninger til DEFF for 2011 indenfor forskningsdata.
2. Projektet har lagt grundlaget til en studierejse til Australien for DEFF-programgruppen for Informationsforsyning.
3. Projektet har som blivende effekt ført til et netværk blandt aktører, samarbejde på konkrete projekter, samt en ny forståelse for behov for kompetenceudvikling på området m.m.
4. Emnet er så omfattende at det skal løses i nationale og internationale partnerskaber.
5. Publikationer :
    a. Heller, Alfred, Nondal, Lars (2011). Forskningsdata Down Under - og måske right here? Revy, 34(1), 16-18. http://www.issuu.com/revy/docs/nummer_1_-_2011_for_web_use
    b. Afslutningsartikel i Revy (ikke udgivet endnu)
6. Projekt Wiki , hvor "Best Practice"-delen kan fremhæves, og hvor en mangfoldighed af ressourcer og rapporteringen af delopgaverne fremgår: se http://forskningsdata.deff.wikispaces.net.
7. Der er gennemgået aktuelle formater, der kan anvendes til forskningsdata: DDI 3.1, DataCite's meget begrænsede og formålstjenlige format, samt det australske format RIS-CS, alle vurderet på wikien.
8. Detaljeret gennemgang af metadata til forskningsdata, DDI og de værktøjer der findes, samt anvendelse på en af prototyperne.
9. En prototypeimplementering af et Fedora-Commons-software med Drupal-frontend til bevaring og visning af forskningsdata på DTU, som vil blive foreslået til koncern-IT i løbet af 2011 som driftskomponent for hele DTU. Dermed føres formentlig projektets arbejde videre som drift.
10. En implementering af DataVerse-softwaren for KUBIS, forventelig i drift i løbet af 2011.

11. DTIC har konkret implementeret software til registrering af DOI under DataCite, som vil kunne danne grundlag for udbredelse af denne identificering af forskningsdata. Ud over softwaren, er der etableret erfaring med anvendelse af DOI for forskningsresultater, og løsningen kan anbefales til en national løsning på området. Resultatet indgår i et nuværende projekt under DEFF-programgruppen for "Arkitektur og Middleware".
12. Projektet, er gennem fælles deltagere, løbende blevet koordineret med aktiviteter i Knowledge Exchange til fælles gavn.
13. Specielt to områder har vist sig at være mere vanskelige end forventet. For det første var det ofte meget svært at få data, der kunne vises offentligt frem, og for det andet var det krævende at komme ind i forskernes processer og etablere de nødvendige tilpasninger for arkivering og langtidsbevaring af data. Opgaverne vil blive videreført af de relevante partnere i projektet, da de er af strategisk vigtighed for organisationerne selv.
14. Der er etableret flere samarbejdsrelationer mellem forskere og biblioteker.
15. Aftalegrundlag mellem forskere og biblioteker eller datacentre, findes fra KE-partnere, DataCite-partnere og vores visit i Australien. Disse aftaler kan overtages og tilpasses til danske forhold. Aftalerne findes på wikien.
16. Formidling af forskningsdata som emne har været på dagsordenen i nationale og internationale fora, hvilket ikke var situationen for få år siden. Projektet har været med til at øge denne omtale gennem aktiv deltagelse ved events og publiceringer.
17. Der er givet eksempler på linkning mellem data og publikationer, hvor DOI-løsningen er blevet de-facto-standarden.
18. Der er etableret en egentlig udstilling af forskningsdata for projektdeltagere, men ikke for den brede offentlighed da rettigheder hertil manglede.
19. Kepler har et konkret eksempel, hvor der er etableret en infrastruktur, hvor forskere, data, projektforslag og preprints er relateret. Alle, som ønsker at arbejde med data, kan få adgang mod at godkende betingelserne for publicering på basis af data. Der arbejdes nu på en langtidsbevaring, herunder bitbevaring og opbygning af ontologier til strukturering.
20. En mulig IT-arkitektur til forskningsdata er dog demonstreret i DTU's konference-proceedings-system under http://conferences.dtu.dk

8 Konklusion

Der hersker ingen tvivl om, at forskningen bliver mere data-tung i de kommende år. Det er heller ingen tvivl om, at de fremkomne data bliver tabt, også i fremtiden, hvis man ikke laver en ekstra indsats på området. Forskningens omdømme kan afhænge af, om der kommer mange eksempler på tabte data i sammenhæng med forskningsfusk. Derfor kan man forudse en øget pres på håndtering, bevaring og tilgængeliggørelse af forskningsdata.

Om der kommer øgede krav om genanvendelse af forskningsdata vides ikke, men også dette synes at være tilfældet og dermed krav om Open Access til forskningsdata.

Mens der for publikationer er udfordringer med Open Access, er der endnu ikke klare tendenser på forskningsdataområdet og man kan nå at holde området "open access" fra starten. Adgangen til dataene vil være differentieret, hvilket en undersøgelse af rettigheder for forskningsdata peger på – rådata kan ikke beskyttes med copyright osv., mens forarbejdede data, under bestemte omstændigheder kan. Rapporten udkommer i løbet af 2011 fra KE.

Projektet viste gennem egne og internationale erfaringer, at forskningsbibliotekerne skulle interessere sig for forskningsdata og håndteringen heraf, som logisk forlængelse af det arbejde de laver med forskningsresultater formidlet i publikationer.

Organisatorisk synes der at være ønsker herom fra de omgivende organisationer, og specielt har universiteter viser øget interesse i emnet. Ledelserne skal dog vurdere deres standpunkt og tilbyde relevante karrieremuligheder, hvis området skal blive bæredygtigt og attraktivt for medarbejdere.

Personalet på forskningsbibliotekerne har stort behov for kvalifikationstilpasninger da forskningsdataområdet virker fremmed for dem. Dette vil skifte gennem indsigt og erfaringer på området, hvorefter opgaven vil indgå lige så naturlig som forskningsregistrering og forskerservices som helhed. Alternativet er at der etableres helt nye uddannelser som "data-stuards, data researchers, data scientists" og andet der er nævnt rundt omkring.

DEFF er et naturligt sted til koordinering af opgraderingen på forskningsdataområde. Dette fremlagte projektet i forhold til strategiarbejdet i DEFF, samt indsats omkring NordForsk og NordBib.

Kilder til erfaringer er speciel turen til Australien og Knowledge Exchange Working Group for hhv. forskningsdata (PRD) og virtuelle forskningsinfrastrukturer (VRE). Da andre lande er langt, langt fremme i forhold til forskningsdata og har investeret trecifret millionbeløb i området, så er det en fornuftig investering at holde sig tæt på de aktive aktører på den internationale scene, bl.a. gennem KE.

Projektet peger på flere muligheder indenfor software, metadata standarder, organisationer osv. der vil kunne være af hjælp til dem der kommer ny ind på

området. Der har ikke vist sig en "entydig" vej endnu og det vil være op til de interesserede, at vælge og vurdere de muligheder der ligger.

Projektet var for kort til at evaluere på Kepler-projektet og automatisk generering af metadata under forskningsprocessen. Internationale projekter viser dog, at det er vejen at gå. Retrospektiv opsamling, som er tilfældet for de opsamlingsaktiviteter i det nærværende projekt, medfører stort besvær med at få informationerne stykket sammen, så resultatet bliver brugbart, selv for simple data som vejrdata, som vi alle sammen forstår.

Det anbefales, at DEFF viderefører de aktiviteter på området som budgettet tillader, for at give biblioteksområdet et forspring i forhold til andre aktører og giver sig selv og sine medlemmer den indsigt i forskningsdata, der er nødvendig for at kunne sikre forskningsresultater og genbrugelighed for fremtiden – dette for at være klar til den tid hvor ministerierne indser, at man skal investere enorme summer på det forsømte område, for at sikre videnskabelig korrekt forskning. Det anbefales at dette gøres forsigtigt her i starten, men øges over tid, da en kritisk masse skal opbygges, for at være attraktiv som arbejdsfelt og som mulig organisatorisk partner på området.

9 Bilag 1: Svar på de mange spørgsmål omkring forskningsdata

Projektet har givet svar på nogle af de mange spørgsmål, der er rejst i de senere år omkring forskningsdata – svarene findes på wikien, men her gives en kort gennemgang uden nærmere diskussion eller dokumentation:

Kan Institutionelle Repositories (IR) anvendes til bevaring af forskningsdata?	På danske universiteter anvendes Pure til forskningsregistrering. Pure er, på nuværende stadie, ikke i stand til at supportere den nødvendige kompleksitet for forskningsdata og kan derfor ikke anbefales som platform for forskningsdata. DataVerse er en open source løsning, der understøtter lagring og deling af forskningsdata. Erfaringer fra DDA viser at installationen, synes at give udfordringer, mens anvendelsen synes at være lige til. Adgangskontrollen er avanceret og versionering er automatisk. Indlæring er nødvendig, men overskuelig. Metadata følger DDI 2, hvilket er særdeles interessant, bl.a. for at afprøve denne standard. DataVerse er installeret hos KB/KUBIS og er ved at blive evalueret. Internationale erfaringer er positive. Fedora-Commons er et open source værktøj, hvorpå der er udviklet nogle overbygninger, som f.eks eSciDoc som kan anvendes på forskningsdata. eSciDoc anvendes af KB til specialer og overvejes til forskningsdata. DTU-softwaren er bygget direkte på Fedora, uden overbygning og har også vist sig at være en fleksibel og anvendelig løsning for forskningsdata (prototype). En endelig brugergrænseflade for forvaltning af data mangler dog endnu.
Kan der etableres det nødvendige samarbejde mellem forskere og biblioteker?	Forskerservice er et aktivitetsfelt for forskningsbiblioteker, som helt naturligt vil lede til arbejde med forskningsdata. Der er altså synergi at spotte.

	Hvis bibliotekerne kan tilbyde en merværdi tyder vores cases, samt den australske erfaring på, at forskerne er åbne for samarbejde. Det kræver dog at bibliotekerne har den nødvendige kompetence og informationsmateriale klar. Samtidigt får bibliotekerne opbakning til forskerservicen.
Aktører og roller på forskningsdataområdet?	Det viser sig på internationalt og nationalt plan, at der er et større uddannelsesbehov for bibliotekarer for at forstå det nye arbejdsfelt - forskningsdata, specielt hvis dataene skal opsamles som en del af en kørende forskningsproces, hvor bibliotekarer tidligere ikke var involveret. Bibliotekerne kan helt klart tilbyde kompetencer, som andre aktører ikke kan byde ind med, speciel informationskompetencer, forståelse for beskrivelser (metadata), spredning af informationer, service-minded personale m.m. På organisatorisk niveau er samarbejdet mellem forskere og biblioteket etableret med stor respekt for bibliotekerne. Usikkerheden ligger i, hvorfor bibliotekerne skulle tage sig af forskningsdata – i Australien viser det sig at dette emne løser sig efter første forsøg. Der skal etableres samarbejde med relevante partnere, herunder universiteternes IT-afdelinger, hvor data ofte ville blive deponeret, datacentre og et kommende nationalt bitarkiv (SA, KB, SB). Bibliotekerne kunne være at tage sig af metadata og retrieval, rettigheder osv. af disse deponeringer.
Hvordan kan bibliotekerne fremstå som relevante og interessante partnere på dataområdet?	Australske erfaringer viser, at hvis biblioteker tager opgaven op, opnår de nødvendige kompetencer og står frem lige så kompetente som de ellers gør, så er de velkomne som hjælp i løsning af denne udfordring.
Hvordan kan der etableres procedurer til indsamling af forsk-	Det vigtigste er at opsamle data mens forskningen foregår, samt dokumentere disse i

ningsdata?	en grad der muliggør genbrug og langtidsbevaring, f.eks. metadata, kalibreringsdata, interviewguides og andet specialviden fra forholdene omkring forskningen.
Kan beskrivelsen skabes automatisk under forskningsprocessen?	Udenlandske projekter viser, at det er nødvendigt for mange forskningsområder, at metadata skabes automatisk under opsamling. Det kan skyldes størrelsen af data eller kompleksiteten af processerne. KB-casen viser, at det er en krævende opgave at skabe sammenhæng mellem data, udstyr m.m.
Hvordan identificeres forskningsdata (DOI-tildeling)?	DOI fungerer fint som identifikator for forskningsdata, specielt med DataCite-metadata og –services. Andre skemaer for ID'er mangler persistens på langt sigt eller er ikke udviklet til det nødvendige niveau. DOI havde den ulempe, at man før sommer 2011 skulle betale et engangsbidrag på 0,04 $ og hvert år en vedligeholdelsesudgift på 0,01 $. Dette var vanskeligt at få etableret. Fra sommer 2011 vil modellen ændres til en medlemskabsmodel, hvor antal DOI ikke påvirker udgifterne. Dette vil øge anvendeligheden for DOI'er også for forskningsdata. DEFF har bevilliget et projekt omkring en national løsning for unikke identifikatorer som i løbet af 2011 vil skabe en prototype og bygger på hhv. DOI og URN. Se Dansk Infrastruktur for Persistente Identifikatorer.
Hvordan kan langtidsbevaring sikres?	Langtidsbevaring af forskningsdata (rå data) ligner langtidsbevaring af andre informationer, dog er kravet for metadata-beskrivelser højere, da data ikke forklarer sig selv – det er ofte kun tal. Herudover skal bitbevaringen sikres, hvilket de nationale infrastrukturer vil være et bud på.

Afrapportering af projektet på det økonomisk-administrative område gennemføres efter aflevering af denne rapport.

[i] Høringssvar fra DEFF om Nordforsks udkast til en nordisk strategi for eScience: "Nordic eScience Action Plan: 10 Concrete Actions to Implement the Nordic eScience Strategy." DEFF Sekretariatet/MCH den 7.5.2009.
[ii] http://www.datacite.org/
[iii] http://www.knowledge-exchange.info/Default.aspx?ID=284